

SEQUENCE LISTING

<110> Ingram, L et al.

<120> METHODS AND COMPOSITIONS FOR SIMULTANEOUS
SACCHARIFICATION AND FERMENTATION

<130> BCI-024CP

<140>

<141>

<150> 60/214,137

<151> 2000-06-26

<150> 60/219,913

<151> 2000-07-21

<160> 17

<170> PatentIn Ver. 2.0

<210> 1

<211> 450

<212> DNA

<213> Zymomonas mobilis

<220>

<223> promoter

<400> 1

cttttttcggc atgagcaacc aacattttca aggtatcatc ctgatgcgca atatcggcac 60
cggttagcca taaccatttt acctgtccgg cggccttaac accttgatca gatggttcgt 120
ggtggttgta ccttgccgaa gggcaccggg aaaaatgttc gcgtcgggtgt ttteggcccg 180
ggcccgaaag ctgaagaagc taaagctgct ggtgcagaag ttgtcggcgc agaagacctg 240
atggaagcca ttcagggcgg cagcattgat ttgatcgtg atgcccttta tactgaaatt 300
gccttgcgct gccataatga agcagcctcc ggtgttttgg cagatttaag cgctgcctga 360
ttttcgtgat cctctagagt ctatgaaatg gagattcatt tatgcctctc tcttattcgg 420
ataaccatcc agtcatccgc aagcttgccc 450

<210> 2

<211> 1509

<212> DNA

<213> Zymomonas mobilis

<220>

<223> expression
vector

<400> 2

gatcaaccgg caatttatcc acggcatcaa attcgatctg tcttttcccg tatcattggc 60
aataccggca ttctgattac aggccgtggt ttgaatgcgg tatgcagttt tgtctatgtc 120

gcatggacat cccagacatt gggattgaac ctgtttgggtg tcatgctttt gattacgact 180
 tttgctaccc tgatttcgga tattaccggt tttcagtcac ggcaaacctt gctgcattac 240
 ggttcaaaag cttttcagga aaaagatttt aaccaatttg atgatgtcct tgccttttgc 300
 atcagagccg attttttttag tgcggcgata ggtatgttgg tagggttagg cggtatcttg 360
 attttaggca cttcaagatt gggatggcct gccgagggtca agccagatgc cttgctttgt 420
 atgctgatta tactttttat gaatatcggc tgggtccaacc gggatgttgc ggctgtgtaa 480
 ccgctttaaa ctggtcacta tttatgagtt tattacgacc tgcgtcagaa ccggagggtg 540
 tggcattggg tattggcttc atatgccttt ggggtatttt ttgtttatat ggtgcctgac 600
 gcaattcacg ctttttgtca cctgtagtta cgctggcatt tatctctttc accaatatac 660
 ggagcgagca tttccgataa gaaaaatatt tcagagaaaa acgcccgttg aagggatgtg 720
 gaaattcact ttaagcgtca gttttaatga aatcctagac tccattttcc agcagggtgg 780
 cacccttgct attggtagct cactgggggc tggggaagcc gctgtctatc gggtcgcgcg 840
 ccagattagt aacggtttat ccaaaccagc acagatgatg atcggctaac atgcatccac 900
 cggcagcacc ggccgtttta tgcttgggat tattgatatg ccgaaaagga tacaacatct 960
 ggaagaaaaa gacgaaggcc ggaataagcg cccattctgc aaaattgtta caacttagtc 1020
 gcgccatcag ggaatgaaaa atcaatccgt ctttttcggc atgagcaacc aacattttca 1080
 aggtatcatc ctgatgcgca atatcggcat cggttagcca taaccatttt acctgtccgg 1140
 cggccttaat accttgatca gatggttcgt ggtgttggtta ccttgccgaa gggcacccgt 1200
 aaaaatgttc gcgtcgggtg tttcgcccgt ggcccgaag ctgaagaagc taaagctgct 1260
 ggtgcagaag ttgtcggcgc agaagacctg atggaagcca ttcagggcgc cagcattgat 1320
 ttogatcgtg atgcccttta tactgaaatt gccttgcgct gccataatga agcagcctcc 1380
 ggtgttttgg cagatttaag cgctgcctga ttttcgtgat cctctagagt ctatgaaatg 1440
 gagattcatt tatgcctctc tcttatcgg ataaccatcc agtcatccgc aagcttggcc 1500
 gtaatccat 1509

<210> 3
 <211> 27
 <212> DNA
 <213> Escherichia coli

<220>
 <223> Description of Artificial Sequence:primer

<400> 3
 cgaattcctg ccgaagtta ttagcca

<210> 4
 <211> 31
 <212> DNA
 <213> Escherichia coli

<220>
 <223> Description of Artificial Sequence:primer

<400> 4
 aaggatcctt ccaccagcta ttgttagtg a 31

<210> 5
 <211> 27
 <212> DNA
 <213> Escherichia coli

<220>
 <223> Description of Artificial Sequence:primer

<400> 5
 agaattctgc cagttggttg acgatag 27

<210> 6
 <211> 30
 <212> DNA
 <213> Escherichia coli

<220>
 <223> Description of Artificial Sequence:primer

<400> 6
 caggatcccc tcaagtcact agttaaactg 30

<210> 7
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:primer

<400> 7
 taatacgact cactataggg 20

<210> 8
 <211> 18
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:primer

<400> 8
 taacaatttc acacagga 18

<210> 9
 <211> 19
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:primer

<400> 9
 cacgacgttg taaaacgac

19

<210> 10
 <211> 30
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:primer

<400> 10
 gactggatgg ttatccgaat aagagagagg

30

<210> 11
 <211> 8
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:primer

<400> 11
 ggcgcgcc

8

<210> 12
 <211> 11544
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:vector

<220>
 <223> all occurrences of n to be sequenced

<220>
 <223> nucleotide positions 1-1451 encodes promoter

<220>
 <223> nucleotide positions 1452-2735 encodes celZ gene

<220>
 <223> nucleotide positions 4916-5776 encodes bla gene

<220>
 <223> nucleotide positions 7061-8251 encodes tet gene

<220>
 <223> nucleotide positions 9476-11544 encodes target sequence from K. oxytoca

```
<220>
<221> CDS
<222> (7061)..(8251)
```

| | | | | | | |
|------------|-------------|-------------|-------------|-------------|-------------|------|
| gatcaaccgg | caatttatcc | acggcatcaa | attcgatctg | tcttttcccg | tatcattggc | 60 |
| aataccggca | ttctgattac | aggccgtggt | ttgaatgcgg | tatgcagttt | tgtctatgtc | 120 |
| gcatggacat | cccagacatt | gggattgaac | ctgttttggtg | tcatgctttt | gattacgact | 180 |
| tttgctaccc | tgatttcgga | tattaccogt | tttcagtcac | ggcaaaccct | gctgcattac | 240 |
| ggttcaaaag | cttttcagga | aaaagatttt | aaccaatttg | atgatgtcct | tgcctttttgc | 300 |
| atcagagccg | attttttttag | tgccggcgata | ggtatgttgg | taggggttagg | cggtatcttg | 360 |
| attttaggca | cttcaagatt | gggatggcct | gccgaggtca | agccagatgc | cttgcttttgt | 420 |
| atgctgatta | tacttttttat | gaatatcggc | tggccaacc | gggatgttgc | ggctgtgtaa | 480 |
| ccgctttaaa | ctggtcacta | tttatgagtt | tattacgacc | tgcgtcagaa | ccggagggttg | 540 |
| tggcattggg | tattggcttc | atatgccttt | ggggatatttt | ttgtttatat | ggtgcctgac | 600 |
| gcaattcacg | ctttttgtca | cctgtagtta | cgctggcatt | tatctctttc | accaatatac | 660 |
| ggagcgagca | tttccgataa | gaaaaatatt | tcagagaaaa | acgcccgttg | aagggatgtg | 720 |
| gaaattcact | ttaagcgtca | gttttaatat | aatcctagac | tccattttcc | agcagggtgg | 780 |
| cacccttgct | attggtagct | cactgggggc | tggggaagcc | gctgtctatc | gggtcgcgcg | 840 |
| ccagattagt | aacggtttat | ccaaaccagc | acagatgatg | atcggctaac | atgcatccac | 900 |
| cggcagcacc | ggccgtttta | tgcttgggat | tattgatatg | ccgaaaagga | tacaacatct | 960 |
| ggaagaaaaa | gacgaaggcc | ggaataagcg | cccattctgc | aaaattgtta | caacttagtc | 1020 |
| gcgccatcag | ggaatgaaaa | atcaatccgt | ctttttcggc | atgagcaacc | aacattttca | 1080 |
| aggtatcatc | ctgatgcgca | atatcggcat | cggtagcca | taaccatttt | acctgtccgg | 1140 |
| cggccttaat | accttgatca | gatggttcgt | ggtgttggtta | ccttgccga | gggcaccggt | 1200 |
| aaaaatgttc | gcgtcgggtg | tttcgcccgt | ggcccgaaag | ctgaagaagc | taaagctgct | 1260 |
| ggtgcagaag | ttgtcggcgc | agaagacctg | atggaagcca | ttcagggcgg | cagcattgat | 1320 |
| ctcgatcgtg | atgcccttta | tactgaaatt | gccttgcgct | gccataatga | agcagcctcc | 1380 |
| ggtgttttgg | cagatttaag | cgctgcctga | ttttcgtgat | cctctagagt | ctatgaaatg | 1440 |

| | | | | | | | | | | | | | | | | |
|--|--|--|--|-----|--|--|--|-----|--|--|--|-----|--|--|--|------|
| gagattcatt t atg cct ctc tct tat tcg gat aac cat cca gtc atc gat | | | | | | | | | | | | | | | | 1490 |
| Met Pro Leu Ser Tyr Ser Asp Asn His Pro Val Ile Asp | | | | | | | | | | | | | | | | |
| 1 | | | | | | | | 5 | | | | 10 | | | | |
| agc caa aaa cac gcc cca cgt aaa aaa ctg ttt cta tct tgt gcc tgt | | | | | | | | | | | | | | | | 1538 |
| Ser Gln Lys His Ala Pro Arg Lys Lys Leu Phe Leu Ser Cys Ala Cys | | | | | | | | | | | | | | | | |
| 15 | | | | 20 | | | | 25 | | | | | | | | |
| tta gga tta agc ctt gcc tgc ctt tcc agt aat gcc tgg gcg agt gtt | | | | | | | | | | | | | | | | 1586 |
| Leu Gly Leu Ser Leu Ala Cys Leu Ser Ser Asn Ala Trp Ala Ser Val | | | | | | | | | | | | | | | | |
| 30 | | | | 35 | | | | 40 | | | | 45 | | | | |
| gag ccg tta tcc gtt agc ggc aat aaa atc tac gca ggt gaa aaa gcc | | | | | | | | | | | | | | | | 1634 |
| Glu Pro Leu Ser Val Ser Gly Asn Lys Ile Tyr Ala Gly Glu Lys Ala | | | | | | | | | | | | | | | | |
| | | | | 50 | | | | 55 | | | | 60 | | | | |
| aaa agt ttt gcc ggc aac agc tta ttc tgg agt aat aat ggt tgg ggt | | | | | | | | | | | | | | | | 1682 |
| Lys Ser Phe Ala Gly Asn Ser Leu Phe Trp Ser Asn Asn Gly Trp Gly | | | | | | | | | | | | | | | | |
| | | | | 65 | | | | 70 | | | | 75 | | | | |
| ggg gaa aaa ttc tac aca gcc gat acc gtt gcg tcg ctg aaa aaa gac | | | | | | | | | | | | | | | | 1730 |
| Gly Glu Lys Phe Tyr Thr Ala Asp Thr Val Ala Ser Leu Lys Lys Asp | | | | | | | | | | | | | | | | |
| 80 | | | | 85 | | | | 90 | | | | | | | | |
| tgg aaa tcc agc att gtt cgc gcc gct atg ggc gtt cag gaa agc ggt | | | | | | | | | | | | | | | | 1778 |
| Trp Lys Ser Ser Ile Val Arg Ala Ala Met Gly Val Gln Glu Ser Gly | | | | | | | | | | | | | | | | |
| 95 | | | | 100 | | | | 105 | | | | | | | | |
| ggt tat ctg cag gac ccg gct ggc aac aag gcc aaa gtt gaa aga gtg | | | | | | | | | | | | | | | | 1826 |
| Gly Tyr Leu Gln Asp Pro Ala Gly Asn Lys Ala Lys Val Glu Arg Val | | | | | | | | | | | | | | | | |
| 110 | | | | 115 | | | | 120 | | | | 125 | | | | |
| gtg gat gcc gca atc gcc aac gat atg tat gtg att att gac tgg cac | | | | | | | | | | | | | | | | 1874 |
| Val Asp Ala Ala Ile Ala Asn Asp Met Tyr Val Ile Ile Asp Trp His | | | | | | | | | | | | | | | | |
| | | | | 130 | | | | 135 | | | | 140 | | | | |
| tca cat tct gca gaa aac aat cgc agt gaa gcc att cgc ttc ttc cag | | | | | | | | | | | | | | | | 1922 |
| Ser His Ser Ala Glu Asn Asn Arg Ser Glu Ala Ile Arg Phe Phe Gln | | | | | | | | | | | | | | | | |
| 145 | | | | 150 | | | | 155 | | | | | | | | |
| gaa atg gcg cgc aaa tat ggc aac aag ccg aat gtc att tat gaa atc | | | | | | | | | | | | | | | | 1970 |
| Glu Met Ala Arg Lys Tyr Gly Asn Lys Pro Asn Val Ile Tyr Glu Ile | | | | | | | | | | | | | | | | |
| 160 | | | | 165 | | | | 170 | | | | | | | | |
| tac aac gag ccg ctt cag gtt tca tgg agc aat acc att aaa cct tat | | | | | | | | | | | | | | | | 2018 |
| Tyr Asn Glu Pro Leu Gln Val Ser Trp Ser Asn Thr Ile Lys Pro Tyr | | | | | | | | | | | | | | | | |
| 175 | | | | 180 | | | | 185 | | | | | | | | |
| gcc gaa gcc gtg att tcc gcc att cgc gcc att gac ccg gat aac ctg | | | | | | | | | | | | | | | | 2066 |
| Ala Glu Ala Val Ile Ser Ala Ile Arg Ala Ile Asp Pro Asp Asn Leu | | | | | | | | | | | | | | | | |
| 190 | | | | 195 | | | | 200 | | | | 205 | | | | |
| att att gtc ggt acg ccc agt tgg tcg caa aac gtt gat gaa gcg tcg | | | | | | | | | | | | | | | | 2114 |
| Ile Ile Val Gly Thr Pro Ser Trp Ser Gln Asn Val Asp Glu Ala Ser | | | | | | | | | | | | | | | | |
| | | | | 210 | | | | 215 | | | | 220 | | | | |
| cgc gat cca atc aac gcc aag aat atc gcc tat acg ctg cat ttc tac | | | | | | | | | | | | | | | | 2162 |
| Arg Asp Pro Ile Asn Ala Lys Asn Ile Ala Tyr Thr Leu His Phe Tyr | | | | | | | | | | | | | | | | |
| 225 | | | | 230 | | | | 235 | | | | | | | | |

| | |
|--|------|
| gcg gga acc cat ggt gag tca tta cgc act aaa gcc cgc cag gcg tta | 2210 |
| Ala Gly Thr His Gly Glu Ser Leu Arg Thr Lys Ala Arg Gln Ala Leu | |
| 240 245 250 | |
| aat aac ggt att gcg ctt ttc gtc acc gag tgg ggc gcc gtt aac gcg | 2258 |
| Asn Asn Gly Ile Ala Leu Phe Val Thr Glu Trp Gly Ala Val Asn Ala | |
| 255 260 265 | |
| gac ggc aat ggc gga gtg aac cag aca gat acc gac gcc tgg gta acg | 2306 |
| Asp Gly Asn Gly Gly Val Asn Gln Thr Asp Thr Asp Ala Trp Val Thr | |
| 270 275 280 285 | |
| ttc atg cgt gac aac aac atc agc aac gca aac tgg gcg tta aat gat | 2354 |
| Phe Met Arg Asp Asn Asn Ile Ser Asn Ala Asn Trp Ala Leu Asn Asp | |
| 290 295 300 | |
| aaa agc gaa ggg gca tca acc tat tat ccg gac tct aaa aac ctg acc | 2402 |
| Lys Ser Glu Gly Ala Ser Thr Tyr Tyr Pro Asp Ser Lys Asn Leu Thr | |
| 305 310 315 | |
| gag tcg ggt aaa ata gta aaa tcg atc att caa agc tgg cca tat aaa | 2450 |
| Glu Ser Gly Lys Ile Val Lys Ser Ile Ile Gln Ser Trp Pro Tyr Lys | |
| 320 325 330 | |
| gcg ggc agc gcc gcc agt aca aca acc gat cag tca acc gat acc acc | 2498 |
| Ala Gly Ser Ala Ala Ser Thr Thr Thr Asp Gln Ser Thr Asp Thr Thr | |
| 335 340 345 | |
| atg gca cca ccg ttg acg aac cga cca caa ccg aca cac cgg caa acc | 2546 |
| Met Ala Pro Pro Leu Thr Asn Arg Pro Gln Pro Thr His Arg Gln Thr | |
| 350 355 360 365 | |
| gct gat tgc tgc aat gcc aac gtt tac ccc aac tgg gtt agc aaa gac | 2594 |
| Ala Asp Cys Cys Asn Ala Asn Val Tyr Pro Asn Trp Val Ser Lys Asp | |
| 370 375 380 | |
| tgg gcg ggc cgg cag cga ctc ata acg aag cag gcc aat cga tcg tct | 2642 |
| Trp Ala Gly Arg Gln Arg Leu Ile Thr Lys Gln Ala Asn Arg Ser Ser | |
| 385 390 395 | |
| aca aag gga acc tgt ata ccg caa act ggt aca ctt cat ccg ttc cgg | 2690 |
| Thr Lys Gly Thr Cys Ile Pro Gln Thr Gly Thr Leu His Pro Phe Arg | |
| 400 405 410 | |
| gca gcg att cct cct ggg cac agg ttg gta gct gta act aat tga | 2735 |
| Ala Ala Ile Pro Pro Gly His Arg Leu Val Ala Val Thr Asn | |
| 415 420 425 | |
| ttaatctttt ccccccaaa ataacagggc tgcgattgca gcctgatacg caacattcca | 2795 |
| ttacttaatt gcgttcaaaa gcgcccaaat ccggtgcgct gccttgtaac taatatgatt | 2855 |
| tctcttttcgt acccgcgтта atcagctttg agttagccga cagacggaac agcgaggttg | 2915 |
| ccggcaacgt gccgtcatta tcacgagata cggtagccag cgaggtgtcc aggctgacga | 2975 |
| atcggaacgc gaagccgctg tccgtatcca tgagttgact cgcataccgca ttactgaccg | 3035 |
| ttgcagaagc agacagagac acgttgttgc ggaagtaatg tttctgtcct gactggacgt | 3095 |
| tgctcccgaa agcataatta atgccgtttt tatatgacgt gttattttatt accgtacgcc | 3155 |

| | | | | | | |
|-------------|------------|-------------|-------------|-------------|-------------|------|
| gccgcgttat | tgttctggtc | aaaacctttg | ctcacgttgc | caaacgcgac | gcaacgggta | 3215 |
| atgcgatgat | tgccgaccgc | tggttcctcc | cagtttgaac | ccgttggcat | tgccggcgaa | 3275 |
| cgcgctnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3335 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3395 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3455 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3515 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3575 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3635 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3695 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3755 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3815 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3875 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3935 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 3995 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 4055 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 4115 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 4175 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 4235 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnngatc | ctctagagtc | 4295 |
| gacctgcagg | aattcactgg | ccgtcgtttt | acaacgtcgt | gactgggaaa | accctggcgt | 4355 |
| taccaacttt | aatcgccttg | cagcacatcc | ccctttcgcc | agctggcgta | atagcgaaga | 4415 |
| ggcccgccacc | gatcgccctt | ccaacagtt | gcgcagcctg | aatggcggaat | ggcgccctgat | 4475 |
| gcggtattttt | ctccttacgc | atctgtgcgg | tatttcacac | cgcataggcg | cgcctatgggt | 4535 |
| gcactctcag | tacaatctgc | tctgatgccg | catagttaag | ccagccccga | caccgcgcaa | 4595 |
| caccgcgtga | cgcgccctga | cgggcttgtc | tgctcccggc | atccgcttac | agacaagctg | 4655 |
| tgaccgtctc | cgggagctgc | atgtgtcaga | ggttttcacc | gtcatcaccg | aaacgcgcga | 4715 |
| gacgaaaggg | cctcgtgata | cgcctatttt | tata'ggttaa | tgtcatgata | ataatggttt | 4775 |
| cttagacgtc | aggtggcact | tttcggggaa | atgtgcgcgg | aaccctatt | tgttttatttt | 4835 |
| tctaaataca | ttcaaatatg | tatccgctca | tgagacaata | accctgataa | atgcttcaat | 4895 |
| aatattgaaa | aaggaagagt | atg agt att | caa cat ttc | cgt gtc gcc | ctt att | 4948 |
| | | Met Ser Ile | Gln His Phe | Arg Val Ala | Leu Ile | |

10

aag ccc tcc cgt atc gta gtt atc tac acg acg ggg agt cag gca act 5716
 Lys Pro Ser Arg Ile Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr
 160

atg gat gaa cga aat aga cag atc gct gag ata ggt gcc tca ctg att 5764
 Met Asp Glu Arg Asn Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile
 170 180

aag cat tgg taa ctgtcagacc aagtttactc atatatactt tagattgatt 5816
 Lys His Trp
 185

taaaacttca tttttaattt aaaaggatct aggtgaagat cctttttgat aatctcatga 5876

ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca 5936

aaggatcttc ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac 5996

caccgctacc agcgggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg 6056

taactggctt cagcagagcg cagataccaa atactgtcct tctagtgtag ccgtagttag 6116

gccaccactt caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac 6176

cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt 6236

taccggataa ggcgagcgg tcgggctgaa cgggggggttc gtgcacacag cccagcttgg 6296

agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc 6356

ttcccgaagg gagaaaggcg gacaggtatc cggttaagcgg cagggtcgga acaggagagc 6416

gcacgagggg gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc 6476

acctctgact tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa 6536

acgccagcaa cgcggccttt ttacgggttcc tggccttttg ctggcctttt gctcacatgt 6596

tctttcctgc gttatcccct gattctgtgg ataaccgtat taccgccttt gaggagctg 6656

ataccgctcg ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggcgc 6716

gccagcgga gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc cgattcatta 6776

atgcagctgg cagcagaggt ttcccgaact gaaagcgggc agtgagcgca acgcaattaa 6836

tgtgagttag ctcaactcatt aggcacccca ggctttacac tttatgcttc cggctcgtat 6896

gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg accatgatta 6956

cgccaagctt gcatgccaat tctcatgttt gacagcttat catcgataag ctttaatgcg 7016

gtagtttatc acagttaa at tgctaacgca gtcaggcacc gtgt atg aaa tct aac 7072
 Met Lys Ser Asn

aat gcg ctc atc gtc atc ctc ggc acc gtc acc ctg gat gct gta ggc 7120
 Asn Ala Leu Ile Val Ile Leu Gly Thr Val Thr Leu Asp Ala Val Gly
 10 20

ata ggc ttg gtt atg ccg gta ctg ccg ggc ctc ttg cgg gat atc gtc 7168

Ile Gly Leu Val Met Pro Val Leu Pro Gly Leu Leu Arg Asp Ile Val
30

cat tcc gac agc atc gcc agt cac tat ggc gtg ctg cta gcg cta tat 7216
His Ser Asp Ser Ile Ala Ser His Tyr Gly Val Leu Leu Ala Leu Tyr
40 50

gcg ttg atg caa ttt cta tgc gca ccc gtt ctc gga gca ctg tcc gac 7264
Ala Leu Met Gln Phe Leu Cys Ala Pro Val Leu Gly Ala Leu Ser Asp
60

cgc ttt ggc cgc cgc cca gtc ctg ctc gct tcg cta ctt gga gcc act 7312
Arg Phe Gly Arg Arg Pro Val Leu Leu Ala Ser Leu Leu Gly Ala Thr
70 80

atc gac tac gcg atc atg gcg acc aca ccc gtc ctg tgg atc ctc tac 7360
Ile Asp Tyr Ala Ile Met Ala Thr Thr Pro Val Leu Trp Ile Leu Tyr
90 100

gcc gga cgc atc gtg gcc ggc atc acc ggc gcc aca ggt gcg gtt gct 7408
Ala Gly Arg Ile Val Ala Gly Ile Thr Gly Ala Thr Gly Ala Val Ala
110

ggc gcc tat atc gcc gac atc acc gat ggg gaa gat cgg gct cgc cac 7456
Gly Ala Tyr Ile Ala Asp Ile Thr Asp Gly Glu Asp Arg Ala Arg His
120 130

ttc ggg ctc atg agc gct tgt ttc ggc gtg ggt atg gtg gca ggc ccc 7504
Phe Gly Leu Met Ser Ala Cys Phe Gly Val Gly Met Val Ala Gly Pro
140

gtg gcc ggg gga ctg ttg ggc gcc atc tcc ttg cat gca cca ttc ctt 7552
Val Ala Gly Gly Leu Leu Gly Ala Ile Ser Leu His Ala Pro Phe Leu
150 160

gcg gcg gcg gtg ctc aac ggc ctc aac cta cta ctg ggc tgc ttc cta 7600
Ala Ala Ala Val Leu Asn Gly Leu Asn Leu Leu Leu Gly Cys Phe Leu
170 180

atg cag gag tcg cat aag gga gag cgt cga ccg atg ccc ttg aga gcc 7648
Met Gln Glu Ser His Lys Gly Glu Arg Arg Pro Met Pro Leu Arg Ala
190

ttc aac cca gtc agc tcc ttc cgg tgg gcg cgg ggc atg act atc gtc 7696
Phe Asn Pro Val Ser Ser Phe Arg Trp Ala Arg Gly Met Thr Ile Val
200 210

gcc gca ctt atg act gtc ttc ttt atc atg caa ctg gta gga cag gtg 7744
Ala Ala Leu Met Thr Val Phe Phe Ile Met Gln Leu Val Gly Gln Val
220

ccg gca gcg ctc tgg gtc att ttc ggc gag gac cgc ttt cgc tgg agc 7792
Pro Ala Ala Leu Trp Val Ile Phe Gly Glu Asp Arg Phe Arg Trp Ser
230 240

gcg acg atg atc ggc ctg tcg ctt gcg gta ttc gga atc ttg cac gcc 7840
Ala Thr Met Ile Gly Leu Ser Leu Ala Val Phe Gly Ile Leu His Ala
250 260

ctc gct caa gcc ttc gtc act ggt ccc gcc acc aaa cgt ttc ggc gag 7888
Leu Ala Gln Ala Phe Val Thr Gly Pro Ala Thr Lys Arg Phe Gly Glu

106190766967744779278407888

```

aag  cag  gcc  att  atc  gcc  ggc  atg  gcg  gcc  gac  gcg  ctg  ggc  tac  gtc      7936
Lys  Gln  Ala  Ile  Ile  Ala  Gly  Met  Ala  Ala  Asp  Ala  Leu  Gly  Tyr  Val
      180                                190

```

ttg ctg gcg ttc gcg acg cga ggc tgg atg gcc ttc ccc att atg att 7984
Leu Leu Ala Phe Ala Thr Arg Gly Trp Met Ala Phe Pro Ile Met Ile
200

ctt ctc gct tcc ggc ggc atc ggg atg ccc gcg ttg cag gcc atg ctg 8032
Leu Leu Ala Ser Gly Gly Ile Gly Met Pro Ala Leu Gln Ala Met Leu
210 220

tcc agg cag gta gat gac gac cat cag gga cag ctt caa gga tcg ctc 8080
Ser Arg Gln Val Asp Asp Asp His Gln Gly Gln Leu Gln Gly Ser Leu
230 240

gcg gct ctt acc agc cta act tcg atc act gga ccg ctg atc gtc acg 8128
Ala Ala Leu Thr Ser Leu Thr Ser Ile Thr Gly Pro Leu Ile Val Thr
250

gcg att tat gcc gcc tcg gcg agc aca tgg aac ggg ttg gca tgg att 8176
Ala Ile Tyr Ala Ala Ser Ala Ser Thr Trp Asn Gly Leu Ala Trp Ile
260 270

gta ggc gcc gcc cta tac ctt gtc tgc ctc ccc gcg ttg cgt cgc ggt 8224
Val Gly Ala Ala Leu Tyr Leu Val Cys Leu Pro Ala Leu Arg Arg Gly
280

gca tgg agc cgg gcc acc tcg acc tga atggaagccg gcggcacctc 8271
Ala Trp Ser Arg Ala Thr Ser Thr
290

gctaacggat tcaccactcc aagaattgga gccaatcaat tcttgcgag aactgtgaat 8331

gcgcaaacca acccttgcca gaacatatcc atcgcgctccg ccatctccag cagccgcacg 8391

cggcgcattct cggggtcgac tctagaggat ccccgcaacg ctgtcagcgc tttccagtta 8451

aacggctcca acgtcgccat aggtaattcc tcgcccggcc atacgatcgg gcaggtgccg 8511

ttggctatcg ccgtcgctg actcatcaca ctatcttccg ctgcatcgcg aagggttttg 8571

accacttctt ccattctctc gtgcgccgga tgccatgctc acgtacgcgg cttatcagat 8631

agtcgggcag gccgtcggtc cagcccaatg aggggaagct ggcgtggagc gatgccagca 8691

cctgctcctc aacaccgtaa tggccggcgg cgaacaggca ttcggcggtta agcgcttcca 8751

gccctttaat catcacgctg cggcacatct tgatagccga cacgctgcc aacgtggttac 8811

caccatagcg ggcgttacat ccaagcgtgg tgagtaattc agcaattgcc tctgcctgtg 8871

gtccccccgt caacagcggc gttcggagtg cccctggggg gaccggcgcc atcaccgcta 8931

catcgacata agcgccgggc ttaaagcatt tggcagcctg acgcttggtc tgcggggcga 8991

cggagttaag gtcaagaaaa tactgcgtgt cgggtcatcag cgggtgcagct tgtgagggcga 9051

catccagggc ggatcccgcg gtgacggtgg aaaatatgag ttcggcacct gtcaacgcgt 9111

| | |
|-------|-----|
| <210> | 16 |
| <211> | 30 |
| <212> | DNA |

$\langle 220 \rangle$

<223> kanamycin-resistance gene product is encoded by
the complement of nucleotides 11621 to 10827

<400> 17

| | | | | | | |
|-------------|-------------|-------------|-------------|------------|------------|------|
| atcgatgaat | tgatccgaag | ttcctattct | ctagaaagta | taggaacttc | gaattgtcga | 60 |
| caagcttgat | ctggcttata | gaaattaata | cgactcacta | tagggagacc | ggaattcccc | 120 |
| tgcaggtcga | ctctagagga | tcannnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 180 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 240 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 300 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 360 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 420 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 480 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 540 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 600 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 660 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 720 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 780 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 840 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 900 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 960 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 1020 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 1080 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 1140 |
| nnnagcgcgt | tcgccggcaa | tgccaacggg | ttcaaactgg | gaggaaccag | cggtcggcaa | 1200 |
| tcatcgcatt | accggttgcg | tcgcgtttgg | caacgtgagc | aaagggtttg | accagaacaa | 1260 |
| taacgcggcg | gcgtacggta | ataaataaca | cgtcataata | aaacggcatt | aattatgctt | 1320 |
| tcgggagcaa | cgtccagtcg | ggacagaaac | attacttccg | caacaacgtg | tctctgtctg | 1380 |
| cttctgcaac | ggtcagtaat | gcggatgcga | gtcaactcat | ggatacggac | agcggcttcc | 1440 |
| gcgtccgatt | cgtcagcctg | gacacctcgc | tggctaccgt | atctcgtgat | aatgacggca | 1500 |
| cggtgccggc | aacctcgctg | ttccgtctgt | cggctaactc | aaagctgatt | aacgcgggta | 1560 |
| cgaaagagaa | atcatattag | ttacaaggca | gcgcaccgga | tttgggcgct | tttgaacgca | 1620 |
| attaagtaat | ggaatgttgc | gtatcaggct | gcaatcgcag | ccctgttatt | ttgggggtga | 1680 |
| aaagattaat | caattagtta | cagctaccaa | cctgtgccca | ggaggaatcg | ctgcccggaa | 1740 |
| cggatgaagt | gtaccagttt | gcggtataca | ggttcccttt | gtagacgac | gattggcctg | 1800 |
| cttcgttatg | agtcgctgcc | ggcccgcgca | gtctttgcta | accagttgg | ggtaaacggt | 1860 |
| ggcattgcag | caatcagcgg | tttgccgggtg | tgtcggttgt | ggtcggttcg | tcaacgggtg | 1920 |
| tgccatgggtg | gtatcggttg | actgatcggt | tgttgactgt | gcggcgctgc | ccgctttata | 1980 |
| tggccagctt | tgaatgatcg | attttactat | tttaccgcag | tcggtcaggt | ttttagagtc | 2040 |
| cggataatag | gttgatgccc | cttcgctttt | atcatttaac | gccagtttg | cggtgctgat | 2100 |
| gttgttgtca | cgcataaacg | ttaccagggc | gtcgggtatc | gtctgggttc | ctccgccatt | 2160 |
| gccgtccgcg | ttaacggcgc | cccactcggg | gacgaaaagc | gcaataccgt | tatttaacgc | 2220 |
| ctggcggggt | ttagtgcgta | atgactcacc | atgggttccc | gcgtagaaat | gcagcgtata | 2280 |
| ggcgatatct | ttggcggtga | ttggatcgcg | cgacgcttca | tcaacgtttt | gcgaccaact | 2340 |
| gggcgtaccg | acaataatca | ggttatccgg | gtcaatggcg | cgaatggcgg | aaatcacggc | 2400 |
| ttcggcataa | ggtttaatgg | tattgctcca | tgaaacctga | agcggctcgt | tgtagatttc | 2460 |
| ataaatgaca | ttcggcttgt | tgccatattt | gcgcgccatt | tcctggaaga | agcgaatggc | 2520 |
| ttcactgcga | ttgttttctg | cagaatgtga | gtgccagtcg | ataatcacat | acatatcggt | 2580 |
| ggcgattgcg | gcataccacca | ctctttcaac | tttggccttg | ttgccagccg | ggtcctgcag | 2640 |
| ataaccaccg | ctttcctgaa | cgcccatagc | ggcgcgaaac | atgctggatt | tccagctttt | 2700 |
| tttcagcgac | gcaacgggat | cggctgtgta | gaatttttcc | ccaccccaac | cattattact | 2760 |
| ccagaataag | ctggtgcccg | caaaactttt | ggctttttca | cctgcgtaga | ttttattgcc | 2820 |
| gctaaccgat | aacgggtcaa | cactcgccca | ggcattactg | gaaaggcagg | caaggcttaa | 2880 |
| tcctaaacag | gcacaagata | gaaacagttt | tttacgtggg | gcgtgttttt | ggctatcgat | 2940 |
| gactggatgg | ttatccgaat | aagagagagg | cataaatgaa | tctccatttc | atagactcta | 3000 |
| gaggatcacg | aaaatcaggc | agcgtctaaa | tctgccaaaa | caccggaggc | tgcttcatta | 3060 |
| tggcagcgca | aggcaatttc | agtataaagg | gcatacacgat | cgaaatcaat | gctgccgccc | 3120 |
| tgaatggctt | ccatcaggtc | ttctgcgcgg | acaacttctg | caccagcagc | tttagcttct | 3180 |
| tcagctttcg | ggccacgggc | gaaaacaccg | acgcgaacat | ttttaccggt | gcccttcggc | 3240 |
| aaggtaacaa | caccacgaac | catctgatca | aggtattaag | gccgcgggac | aggtaaaatg | 3300 |
| gttatggcta | accgatgccg | atattgcgca | tcaggatgat | accttgaaaa | tggttggttg | 3360 |
| tcatgccgaa | aaagacggat | tgatttttca | ttccctgatg | gcgcgactaa | gttgtaacaa | 3420 |

0363625201051901

| | | | | | | |
|-----------------------|---------------------|---------------------|-----------------------|---------------------|---------------------|------|
| t t t t g c a g a a | t g g g c g c t t a | t t c c g g c c t t | c g t c t t t t t t c | t t c c a g a t g t | t g t a t c c t t t | 3480 |
| t c g g c a t a t c | a a t a a t c c c a | a g c a t a a a a c | g g c c g g t g c t | g c c g g t g g a t | g c a t g t t a g c | 3540 |
| c g a t c a t c a t | c t g t g c t g g t | t t g g a t a a a c | c g t t a c t a a t | c t g g c g c g c g | a c c c g a t a g a | 3600 |
| c a g c g g c t t c | c c c a g c c c c c | a g t g a g c t a c | c a a t a g c a a g | g g t g c c a c c c | t g c t g g a a a a | 3660 |
| t g g a g t c t a g | g a t t t c a t t a | a a a c t g a c g c | t t a a a g t g a a | t t t c c a c a t c | c c t t c a a c g g | 3720 |
| g c g t t t t t t c t | c t g a a a t a t t | t t t c t t a t c g | g a a a t g c t c g | c t c c g t a t a t | t g g t g a a a g a | 3780 |
| g a t a a a t g c c | a g c g t a a c t a | c a g g t g a c a a | a a a g c g t g a a | t t g c g t c a g g | c a c c a t a t a a | 3840 |
| a c a a a a a a t a | c c c c a a a g g c | a t a t g a a g c c | a a t a a c c a a t | g c c a c a a c c t | c c g g t t c t g a | 3900 |
| c g c a g g t c g t | a a t a a a c t c a | t a a a t a g t g a | c c a g t t t a a a | g c g g t t a c a c | a g c c g c a a c a | 3960 |
| t c c c g g t t g g | a c c a g c c g a t | a t t c a t a a a a | a g t a t a a t c a | g c a t a c a a a g | c a a g g c a t c t | 4020 |
| g g c t t g a c c t | c g g c a g g c c a | t c c c a a t c t t | g a a g t g c c t a | a a a t c a a g a t | a c c g c c t a a c | 4080 |
| c c t a c c a a c a | t a c c t a t c g c | c g c a c t a a a a | a a a t c g g c t c | t g a t g c a a a a | g g c a a g g a c a | 4140 |
| t c a t c a a a t t | g g t t a a a a t c | t t t t t c c t g a | a a a g c t t t t g | a a c c g t a a t g | c a g c a a g g t t | 4200 |
| t g c c a t g a c t | g a a a a c g g g t | a a t a t c c g a a | a t c a g g g t a g | c a a a a g t c g t | a a t c a a a a g c | 4260 |
| a t g a c a c c a a | a c a g g t t c a a | t c c c a a t g t c | t g g g a t g t c c | a t g c g a c a t a | g a c a a a a c t g | 4320 |
| c a t a c c g c a t | t c a a a a c a c g | g c c t g t a a t c | a g a a t g c c g g | t a t t g c c a a t | g a t a c g g g a a | 4380 |
| a a g a c a g a t c | g a a t t t g a t g | c c g t g g a t a a | a t t g c c g g t t | g a t c g a t c c c | c g g g t a c c g a | 4440 |
| g c t c g a a t t c | c g a g c t t g g c | g c g c c t a t g c | g g t g t g a a a t | a c c g c a c a g a | t g c g t a a g g a | 4500 |
| g a a a a t a c c g | c a t c a g g c g c | c a t t c g c c a t | t c a g g c t g c g | c a a c t g t t g g | g a a g g g c g a t | 4560 |
| c g g t g c g g g c | c t c t t c g c t a | t t a c g c c a g c | t g g c g a a a g g | g g g a t g t g c t | g c a a g g c g a t | 4620 |
| t a a g t t g g g t | a a c g c c a g g g | t t t t c c c a g t | c a c g a c g t t g | t a a a a c g a c g | g c c a g t g a a t | 4680 |
| t c c a t c a a c g | c t t g c t g t a a | c c a g g a g c c a | a a g c t a t g a a | t g t a c c t t t t | a g c t a c t c g t | 4740 |
| c a c c c a c c c t | g a g c g t t g a g | g c g t t a a a g c | a c t c t a t t g c | t t a t a a g c t g | a t g t t t a t c a | 4800 |
| t c g g c a a a g a | c c c g g c t a t c | g c t a a c a a g c | a t g a a t g g c t | c a a c g c c a c g | c t g t t c g c c g | 4860 |
| t t c g c g a t c g | t a t g g t t g a g | c g c t g g c t g c | g c t c a a a c c g | c g c g c a c g t c | t c t c a g g a a g | 4920 |
| t t c g c c a g g t | t t a c t a c c t g | t c g a t g g a a t | t t t t g a t t g g | c c g t a c g t t g | t c c a a c g c g c | 4980 |
| t g t t a t c g c t | c g g c a t t t a t | g a g g a t g t g a | a c a g c g c g c t | g g a a g a g a t g | g g g c t g a a c c | 5040 |
| t t g a a g a a t t | a a t t g a t g a a | g a a a a c g a c c | c g g g c t t a g g | c a a c g g c g g t | c t t g g t c g t c | 5100 |
| t g g c g g c c t g | c t t c c t c g a t | t c g c t t g c g g | c g c t g g g g t t | a c c g g g c c g c | g g c t a c g g t a | 5160 |
| t t c g c t a c g a | c t a c g g g a t g | t t t a a g c a g a | a t a t c g t c g a | t g g g c g g c a g | a a a g a a t c c c | 5220 |
| c g g a t t a c t g | g c t g g a a t a c | g g t a a c c c g t | g g g a g t t c g a | g c g c c a t a a t | a c g c g c t a c a | 5280 |
| a a g t g c g c t t | c g g c g g a c g c | a t t c a g c a g g | a a g g t a a a t a | c t c c c g c t g g | g t g g a g a c c g | 5340 |
| a a g a g a t t a t | t g c c g a a g c c | t a t g a c c a g a | t t a t c c c t g g | c t t c g a c a c c | g a c g c c a c c a | 5400 |
| a c a c g c t g c g | c c t g t g g a g c | g c c c a g g c c a | g c a g c g a g a t | t a a c c t c g g t | a a a t t c a a c c | 5460 |
| a g g g c g a c t a | c t t c g c g g c g | g t g g a a g a t a | a a a a c c a t t c | c g a g a a c g t g | t c g c g g g t a c | 5520 |
| t c t a t c c g g a | t g a c t c g a c c | t a t t c a g g a c | g c g a g c t g c g | c c t g c g g c a g | g a g t a c t t c c | 5580 |
| t c g t t t c g g c | g a c g g t g c a g | g a c a t c c t c a | g c c g c c a c t a | c c a g c t g c a t | a a a a c c t a c g | 5640 |
| c c a a c c t g g c | g g a c a a a a t c | g c g a t t c a t c | t c a a c g a c a c | g a a c c c g g t g | c t g t c g a t t c | 5700 |
| c g g a g c t g a t | g c g c c t g c t g | a t t g a c g a g c | a t a a g a t c a g | c t g g g a t g a g | g g n n n n n n n n | 5760 |
| n n n n n n n n n n | n n n n n n n n n n | n n n n n n n n n n | n n n n n n n n n n | n n n n n n n n n n | n n n n n n n n n n | 5820 |
| n n n n n n n n n n | n n n n n n n n n n | n n n n n n n n n n | n n n n n n n n n n | n n n n n n n n n n | n n n n n n n n n n | 5880 |
| n n n n n n n n n n | n n n n n n n n n n | n n n n n n n n n n | n n n n n n n n n n | | | |

| | | | | | | |
|-------------|-------------|-------------|------------|-------------|------------|-------|
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 7140 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 7200 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 7260 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 7320 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 7380 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 7440 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 7500 |
| nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | nnnnnnnnnn | 7560 |
| nnnnnnngaa | ttcgagctcg | gtacccgggg | atcgacataa | ccgataggtc | ctgcattgat | 7620 |
| ggactgaaag | gtttcgacgg | cggttgtgtg | ggttttgctt | gcccatctgg | cggcatgaat | 7680 |
| agtgtcattc | atgacgatcc | agttcgatat | tcaacagacc | gtctttgtaa | tcggcaccga | 7740 |
| caattttgat | caataaagcg | tttgacctga | tgcattgagg | taaatccatt | cgttcggttg | 7800 |
| ttcttttctg | attacctgtc | ctgttaacct | gtggatatag | aaggtcggtt | caatgagtag | 7860 |
| tattctgacg | catctgacaa | ttggttccaa | tgacctgaag | aaggcgcgca | tcttttatga | 7920 |
| tgctgttttg | gaaccgttgg | gtatcaaact | tattcgcgag | gtcgaaggac | agcgttttgc | 7980 |
| ctatggtaaa | gacggcgaag | aaggacgcat | catcattgta | aagcctatta | atggtgaagc | 8040 |
| cgctaccgct | ggaaatggta | tcactatcgg | tttggcagcg | ccttctgatg | aagctgtcga | 8100 |
| tgctttttat | aaagcaggct | tggctaattg | cggtaaggat | gccggagaac | cggggcctcg | 8160 |
| tccggtgctg | aataattctc | ggggtgccta | tttatatgac | cctgaaggca | ataaaatctg | 8220 |
| cgctttcaat | tttaaataag | atttcttttg | tgcagggtta | ttcaaaatag | ccctgcattt | 8280 |
| tcagtattat | agcggccatt | atggcttttg | ccttgataaa | aaatttatca | gggctgtttt | 8340 |
| tcgtgatgaa | tattttttgat | ttttcaagaa | aagcctgata | tcttccaaca | tctttccttg | 8400 |
| tatataaatg | gagcgagcta | tggcgcgcgt | aactgtcgaa | gactgtatcg | ataaagttca | 8460 |
| taatcgtttc | gatttgatcc | tctagagtca | acctgcttgt | tactcgtgat | cccattcaca | 8520 |
| agggcgaatt | aattcgccct | tctgttccgt | taccaacact | tgagccggag | gcataatggg | 8580 |
| aaagccaatg | tggcgttggt | gggcgttgat | gctgatgggt | tggttcagtg | cgtcggctac | 8640 |
| ggcggcgaac | ggctgggaaa | tctataaaaag | ccgtttcatg | accacggacg | ggcgcattca | 8700 |
| ggataccggc | aataagaatg | tcagccacac | cgaaggtcag | ggattcgcca | tgctgatggc | 8760 |
| ggtgcattac | gatgaccgca | tcgcgttcga | taacctgtgg | aactggacgc | aaagccacct | 8820 |
| gcggaacacg | accagcggct | tgttctactg | gcgttacgat | ccgtcggcgg | ccaatccggt | 8880 |
| ggtggataag | aacaacgcct | cggatggcga | tgtgctgatt | gcctgggcgt | tgttaaaagc | 8940 |
| gggaaataag | tggcaggaca | accgttacct | gcaggcgtcg | gacagcatcc | agaaagcgat | 9000 |
| catcgccagc | aatatcattc | agtttgcggg | ccgcaccgtg | atgttgcccg | gcgcctatgg | 9060 |
| tttcaacaag | aacagctatg | tgatccttaa | cccgtcgtat | ttcctgttcc | cggcctggcg | 9120 |
| cgactttgct | aaccgcagcc | atcttcagggt | gtggcggcaa | ctgattgacg | acagcctgtc | 9180 |
| attggtcggg | gaaatgcggt | tcggtcagggt | cgggctgccg | acggactggg | cggcgtgaa | 9240 |
| cgcggatggc | tcgatggcgc | cggcgacggc | ctggccgtcg | cgtttcagtt | acgacgccat | 9300 |
| tcgtatcccg | ctgtaatttg | actggtatga | cgccaaaacc | acggcgctgg | tgccgttcca | 9360 |
| gctgtactgg | cgtaactatc | cccgcctgac | gacgcgggcc | tgggttgatg | tgctgagcag | 9420 |
| taacaccgcg | acttacaata | tgcagggcgg | tttgctggcg | gtgcgcgacc | tgacgatggg | 9480 |
| caacctcgac | gggctcagcg | atctgccagg | cgcacgcgaa | gattactact | cgtcgagcct | 9540 |
| gcgcctgctg | gtgatgttgg | cgcgcggtaa | ataaccttat | tcttgcggtg | cacatggcga | 9600 |
| ggacgatgtc | cttgccattt | tccccacttt | tatccctctg | aatggcgtgt | ttttcacgct | 9660 |
| ttgttaacct | gcttgttact | cgtgatccca | ttcacaaggg | cgaattgacc | tgcaggcatg | 9720 |
| caagcttggc | gtaatcatgg | tcatagctgt | ttcctgtgtg | aaattgttat | ccgctcacia | 9780 |
| ttccacacaa | catacgagcc | ggaagcataa | agtgtaaagc | ctgggggtgcc | taatgagtga | 9840 |
| gctaactcac | attaattgcg | ttgcgctcac | tgcccgtttt | ccagtcggga | aacctgtcgt | 9900 |
| gccagctgca | ttaatgaatc | ggccaacgcg | cggggagagg | cggtttgctg | attgggcgct | 9960 |
| cttccgctgg | cgcgccagggt | cgactctaga | ggatccccgg | ggaagatctt | ccggaagatc | 10020 |
| ttcccagagct | cgaatttaatt | ccgcgatgaa | ttgatccccg | aagttcctat | tctctagaaa | 10080 |
| gtataggaac | tcgaattgggt | cgacaagcta | gcttgcatgc | aagcttgtag | tctatagtgt | 10140 |
| cacctaaatc | gtatgtgtat | gatacataag | gttatgtatt | aattgtagcc | gcgttctaac | 10200 |
| gacaatatgt | acaagcctaa | ttgtgtagca | tctggcttac | tgaagcagac | cctatcatct | 10260 |
| ctctcgtaaa | ctgccgtcag | agtcggtttg | gttggacgaa | ccttctgagt | ttctggtaac | 10320 |
| gccgttccgc | accccggaaa | tggtcagcga | accaatcagc | agggtcacgc | ctagcccatg | 10380 |
| gctaattctg | tcagccgtta | agtgttcctg | tgtcactgaa | aattgctttg | agaggctcta | 10440 |
| agggcttctc | agtgcgttac | atccctggct | tgttgtccac | aaccgttaaa | ccttaaaagc | 10 |

| | | | | | | |
|------------|-------------|------------|-------------|------------|------------|-------|
| cgtactatca | acaggttga | ctgcggatct | tgcggccgca | aaaattaaaa | atgaagtttt | 10800 |
| gacggtatcg | aaccccagag | tcccgtcag | aagaactcgt | caagaaggcg | atagaaggcg | 10860 |
| atgcgctgcg | aatcgggagc | ggcgataccg | taaagcacga | ggaagcggtc | agcccattcg | 10920 |
| ccgccaagct | cttcagcaat | atcacgggta | gccaacgcta | tgtcctgata | gcggtcgcc | 10980 |
| acacccagcc | ggccacagtc | gatgaatcca | gaaaagcggc | cattttccac | catgatattc | 11040 |
| ggcaagcagg | catcgccatg | ggtcacgacg | agatcctcgc | cgtcgggcat | ccgcgccttg | 11100 |
| agcctggcga | acagttcggc | tggcgcgagc | ccctgatgct | cttcgtccag | atcatcctga | 11160 |
| tcgacaagac | cggttccat | ccgagtacgt | gctcgctcga | tgcgatgttt | cgcttggtgg | 11220 |
| tcgaatgggc | aggtagccgg | atcaagcgta | tgcagccgcc | gcattgcata | agccatgatg | 11280 |
| gatactttct | cggcaggagc | aaggtgagat | gacaggagat | cctgccccgg | cacttcgccc | 11340 |
| aatagcagcc | agtcctttcc | cgcttcagtg | acaacgtcga | gcacagctgc | gcaaggaacg | 11400 |
| cccgtcgtgg | ccagccacga | tagccgcgct | gcctcgtctt | ggagttcatt | cagggcaccg | 11460 |
| gacaggtcgg | tcttgacaaa | aagaaccggg | cgccccctgcg | ctgacagccg | gaacacggcg | 11520 |
| gcatcagagc | agccgattgt | ctgttgtgcc | cagtcatagc | cgaatagcct | ctccacccaa | 11580 |
| gcggccggag | aacctgcgtg | caatccatct | tgttcaatca | tgcgaaacga | tcctcatcct | 11640 |
| gtctcttgat | ccactagatt | attgaagcat | ttatcagggt | tattgtctca | tgagcggata | 11700 |
| catatttgaa | tgtattttaga | aaaataaaca | aataggggtt | ccgcgcacat | ttccccgaaa | 11760 |
| agtgccacct | gc | | | | | 11772 |

10800 10860 10920 10980 11040 11100 11160 11220 11280 11340 11400 11460 11520 11580 11640 11700 11760 11772